

AN ANALYSIS OF RISK REDUCTION CHOICES IN DCIS BREAST CANCER PATIENTS

A Senior Project
presented to
the Faculty of the Statistics Department
California Polytechnic State University, San Luis Obispo

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Science

by
Lauren Soltesz
December 2012
Advisor: Rebecca Ottesen

© 2012 Lauren Soltesz

Abstract

The main focus of this paper was to evaluate possible demographic and clinical characteristics associated with a woman's choice of breast conserving surgery (BCS), unilateral mastectomy (ULM), or bilateral risk reduction mastectomy (BRRM). The cohort consisted of patients presenting to the City of Hope National Medical Center with ductal carcinoma in situ breast cancer who elected to have cancer directed surgery (N=305). Analyses to examine associations of patient characteristics with type of surgery were conducted using a multinomial logistic regression. Results showed that older women were more likely to choose breast conserving surgery over bilateral risk reduction mastectomy than younger women (OR=6.64, 95% CI=2.02-21.84). Women diagnosed between 1997 and 2004 were more likely to choose BCS over BRRM than patients more recently (2005-2012), (OR=3.91, 95% CI=1.48-10.33). Women with small tumors (<2cm) and multicentric disease, were also more likely to choose BCS over BRRM than women with large tumors 2cm or more (OR=3.35, 95% CI=1.24-9.05) and those without multicentric disease (OR=6.69, 95% CI=2.44, 18.32). Those with a positive hormone receptor status were more likely to choose BCS over BRRM than those with an unknown hormone receptor status (OR=5.28, 95% CI=1.99-13.96), while there was no association found between negative and positive hormone receptor status and surgical choice. Comparisons were not significant when examining likeliness of choosing unilateral mastectomy over BRRM. Binomial logistic regression was conducted to investigate likeliness of undergoing hormone therapy for those patients who chose not to get BRRM, and the cohort was reduced to patients with at least 270 days follow-up (N=243). Results found women with positive or unknown hormone receptor status were substantially more likely to undergo hormone therapy than women with a negative hormone receptor status (OR=15.22, 95% CI=3.25-71.24, and OR=12.07, 95% CI=2.60-56.05 respectively). The interaction between age at diagnosis and radiation therapy use also showed that for a woman who has not received radiation therapy, the estimated odds of undergoing hormone therapy were 3.35 times higher for those between 50 and 64 years of age compared to those younger than 50 years (95% CI=1.25-8.99). Finally survival analysis was conducted to examine if BRRM impacts the chances of survival in women with DCIS breast cancer. The proportion of deaths in this disease subset is small (2.3%), and there wasn't enough evidence to conclude a significant difference in survival based on surgery type.

Table of Contents

Introduction	4
Methods	5
Data	5
Variables of Interest	6
Statistical Analysis	7
Results	10
Surgical Choice	10
<i>Descriptive Statistics</i>	10
<i>Modeling</i>	14
Hormone Therapy	15
<i>Descriptive Statics</i>	15
<i>Modeling</i>	20
Survival Analysis	21
Conclusion	22
Further Research	25
References	26
Appendix	27

Introduction

Ductal Carcinoma In Situ (DCIS) is a form of breast cancer associated with a higher risk for subsequent contralateral breast cancer. The main choice of treatment is surgery with or without adjuvant radiation therapy and/or hormone therapy. One of the more aggressive measures for preventing contralateral breast cancer is for a woman to get bilateral risk reduction mastectomy, which consists of removing the breast with the tumor as well as the unaffected breast. When a woman has both breasts removed she doesn't need to go on to get further treatment. More and more patients seem to be choosing this procedure and doctors would like to investigate what characteristics of a patient lead to this choice. This project examined certain demographic and clinical characteristics of City of Hope breast cancer patients to see if the likeliness of a patient choosing a bilateral mastectomy can be predicted. There were three main research questions addressed in this project. The first is what factors influenced the risk reduction choice a patient made, between breast conserving surgery (lumpectomy), unilateral mastectomy, or bilateral risk reduction mastectomy. The second research question is of those women who choose not to get a bilateral risk reduction mastectomy, what factors predict their choice of whether or not they had hormone therapy. The final question of interest is whether surgical choice impacted a woman's survival outcome.

Methods

Data

The data used for this analysis comes from breast cancer patients presenting to the City of Hope National Medical Center between July 1997 and February 2012. Only patients with a breast cancer histology of DCIS for the episode of interest, at least 90 days of good quality follow-up data post presentation, an overall programmed stage 0 at diagnosis, and who had a definitive surgery were included (N=305). The timing rule of 90 days post diagnosis was chosen to ensure adequate time to allow for definitive surgery. Additionally for the hormone therapy analysis, this cohort was further reduced to only patients with at least 270 days of good quality follow-up data post presentation and excluded patients with bilateral risk reduction mastectomy (N=243). More time was allowed for the hormone therapy analysis to allow for time to initiate adjuvant hormone therapy post-surgery.

The data tables came from a relational database model and were accessed using SAS. There were raw data sets that contained the data entered directly into the database, as well as derived data sets that were manipulations of the raw data using predefined algorithms. Each data table had a data dictionary used to explain variable meanings and coding, as well as pre-defined SAS formats. Additional formats were created for use in the analysis to collapse and categorize the variables of interest.

A large part of this project was preparing the data sets for final analysis. The data needed to be narrowed down to include only a subset of patients and only certain variables of interest from the various data sets. Some aspects of data preparation involved detailed programming and data manipulation. For example women who had bilateral risk reduction mastectomies have two mastectomy records stored as different rows in the data table. This information was then combined with the laterality of the cancer to compare the side of the surgery in order to create the surgical choice variable.

Variables of Interest

Initially ten predictors of interest that relate to treatment choice were defined. Age was defined as the woman's age at diagnosis with breast cancer, and was categorized based on meaningful clinical specifications. Race was derived from the woman's racial background and Spanish/Hispanic ethnicity. Education referred to the highest level of education completed at time of diagnosis. Year of diagnosis was categorized as 1997-2004, or 2005-2012 and included as a predictor to see if the trend of bilateral risk reduction mastectomies was a more recent occurrence. Histological grade referred to a system that is used to classify cancer cells in terms of how abnormal they look under a microscope and how quickly the tumor is likely to grow and spread. This can be an important factor when developing a treatment plan and was therefore a variable of consideration in predicting surgical choice. A woman's type of health insurance was also suspected to impact treatment decisions. Tumor size was measured as the size of mass in centimeters as established by surgery, which was under consideration to investigate if larger tumors impacted treatment decisions. This variable is categorized as either less than 2cm, 2cm or more, or unknown, as doctors consider 2cm to be the cutoff between small and large tumors. Multicentric disease involves two or more distinct primary tumors, usually in different quadrants of the breast. Menopausal status was of interest in regards to whether a woman is pre or post menopause. Hormone receptor status was included as a combination of the patient's estrogen and progesterone receptor status at diagnosis. About 75% of women with breast cancer have tumors that contain estrogen receptors (called ER-positive). About 65% of these are also PR-positive, meaning they grow in response to the hormone progesterone. Women who are ER/PR positive are more likely to respond to hormone treatment than women who are ER/PR negative and therefore may not need to consider aggressive surgery as an option.

For the hormone therapy analysis, all of the previously described variables were also of interest. In addition, the surgical choice, BCS or unilateral mastectomy, was also of interest. Surgical margins

were determined during or after surgery when a pathologist examines the rim of tissue to be sure it's clear of any cancer cells. If the surgical margins are negative, then no cancer cells were seen at the outer edge of the tissue that was removed. However if the margins are positive, cancer cells came right out to the edge of the removed tissue and this might have been a factor in predicting a woman's likeliness to undergo hormone therapy. The radiation therapy variable took into consideration whether or not a patient has received radiation therapy which uses high-energy radiation to kill cancer cells by damaging their DNA.

All of the variables under consideration for analysis of surgical choice were also considered potential adjustment variables for survival analysis (age, race, education, year, grade, insurance, tumor size, multicentric disease, menopausal status, and hormone receptor status).

Statistical Analysis

Descriptive statistics were obtained in SAS to gain a better understanding of patterns in the data based on surgical choice. For hormone therapy the frequencies and percentages of each variable categorized by whether or not the patient received hormone therapy were also reviewed. These descriptives also contained the univariate p-values for each variable obtained using logistic regression, based on the relevant outcome.

Multinomial logistic regression was used to predict the likeliness of a patient choosing breast conserving surgery, unilateral mastectomy, or bilateral risk reduction mastectomy. Multinomial logistic regression was run using PROC LOGISTIC in SAS and the link=glogit option. This option fit the generalized logit function where each nonreference response category was contrasted with the reference category. Bilateral Risk Reduction Mastectomy (BRRM) was used as the reference response category as it is the surgical choice of interest for comparison. The saturated model included potential predictors with univariate p-values less than 0.2: age, education, year, insurance, tumor size, multicentric disease, menopausal status, and hormone receptor status. From this a predictor with the largest p-value was

eliminated from the model until all remaining predictors were significant at the 0.05 level. A priori interactions with age were also tested and only included if they were found be useful contributors to the model. Multinomial logistic regression produced two different regression equations. The first referred to the likeliness of a woman choosing breast conserving surgery over a bilateral risk reduction mastectomy. The second referred to the likeliness of a woman choosing a unilateral mastectomy over a bilateral risk reduction mastectomy. Results included the type 3 analysis of effects p-values, odds ratio estimates and 95% confidence intervals for the odds ratios for each of the comparisons to BRRM.

Logistic regression assumptions were checked for the final model. The first assumption was that there is linearity between the logits and explanatory variables. This assumption does not apply for this analysis however because the explanatory variables were categorical. The independent observations assumption was also satisfied because we can assume one patient's experience with breast cancer has nothing to do with another patient's experience. There were 305 patients in the final analysis data set so the decent sample size assumption was met. Explanatory variables were checked for multicollinearity using PROC CORR and spearman correlation coefficients were calculated. None of the correlations appeared to be large, indicating that multicollinearity was not an issue with this model. Deviance and Pearson goodness of fit statistics for the multinomial logistic regression residuals were tested. Deviance produced a ratio of value to degrees of freedom of 0.8499 and a p-value of 0.8896. Pearson produced a ratio of 0.7898 and a p-value of 0.962. Since both of these ratios were close to 1, there was no indication of sever under-dispersion or over-dispersion.

Binary logistic regression was carried out to model the probability of undergoing hormone therapy, yes or no. A similar approach to the surgery analysis was employed considering potential predictors to be univariate p-values that were less than 0.2 and then eliminating predictors with the largest p-value one at a time. Again a priori interactions with age were also tested. Results included the type 3 analysis of effects p-values, odds ratio estimates and 95% confidence intervals. Assumptions

were checked using the lackfit option in PROC LOGISTIC which produced the Hosmer and Lemeshow Goodness-of-fit test. The chi-square test statistic is 9.7475 with 9 degrees of freedom, and the p-value was 0.3713 which indicated an adequate fit. Deviance and Pearson goodness of fit statistics for the model residuals were tested. Deviance produced a ratio of value to degrees of freedom of 1.4082 and a p-value of 0.1693. Pearson produced a ratio of 1.3286 and a p-value of 0.2081. Since both of these ratios were close to 1, there was no indication of severe under-dispersion or over-dispersion. Residual diagnostic plots were produced to examine the standardized Pearson residual values, cook's distance, and leverage. No observations were found to have an extreme impact on the fit of the model.

Initial survival analyses were conducted to model time until death from breast cancer. DCIS has a relatively low death rate and in this cohort there were only 7 deaths (2.3%). PROC LIFETEST was used stratified on type of surgery, and the logrank test was used to compare strata.

Results

Surgical Choice

Descriptive Statistics

By examining the frequency of surgical choice, we saw that bilateral risk reduction mastectomy appeared to be used more widely in women who were <50 years of age (61.5%) while breast conserving surgery was used more in women over 65 years (28.7%) compared to 20.5% for unilateral mastectomy and 10% for BRRM. Age at diagnosis, education, year, insurance, tumor size, multicentric disease, menopausal status, and hormone receptor status were potential predictors for the model ($p < .2$). These were factors likely to be associated with a woman's surgical choice, for example, the likeliness of a particular surgical choice could vary depending on how old she is or what year she was diagnosed.

Table 1a: Demographics and Clinical Characteristics of Women with Stage 0 DCIS Disease by Surgical Choice

		Surgical Choice				P-Value ₁
		ALL (N=305)	Breast Conserving Surgery (N=178)	Unilateral Mastectomy (N=88)	Bilateral Risk Reduction Mastectomy (N=39)	
Age at Diagnosis						
<50	N	113	52	37	24	0.0037
	Col (%)	37.05	29.21	42.05	61.54	
50-64	N	119	75	33	11	
	Col (%)	39.02	42.13	37.50	28.21	
>=65	N	73	51	18	4	
	Col (%)	23.93	28.65	20.45	10.26	

		Surgical Choice				P-Value ₁
		ALL (N=305)	Breast Conserving Surgery (N=178)	Unilateral Mastectomy (N=88)	Bilateral Risk Reduction Mastectomy (N=39)	
Race						
Caucasian	N	164	102	41	21	0.3448
	Col (%)	53.77	57.30	46.59	53.85	
Hispanic	N	67	40	20	7	
	Col (%)	21.97	22.47	22.73	17.95	
Other	N	74	36	27	11	
	Col (%)	24.26	20.22	30.68	28.21	
Education						
High School or Less	N	139	78	49	12	0.0283
	Col (%)	45.57	43.82	55.68	30.77	
More Than High School	N	166	100	39	27	
	Col (%)	54.43	56.18	44.32	69.23	
Insurance						
Other	N	4	2	0	2	0.0758
	Col (%)	1.31	1.12	0	5.13	
Managed	N	172	93	50	29	
	Col (%)	56.39	52.25	56.82	74.36	
Medicare	N	81	56	21	4	
	Col (%)	26.56	31.46	23.86	10.26	
Medicaid	N	48	27	17	4	
	Col (%)	15.74	15.17	19.32	10.26	

		Surgical Choice				P-Value ₁
		ALL (N=305)	Breast Conserving Surgery (N=178)	Unilateral Mastectomy (N=88)	Bilateral Risk Reduction Mastectomy (N=39)	
Menopausal Status						
Pre	N	125	57	42	26	0.0012
	Col (%)	40.98	32.02	47.73	66.67	
Post	N	179	121	46	12	
	Col (%)	58.69	67.98	52.27	30.77	
Conflict	N	1	0	0	1	
	Col (%)	0.33	0	0	2.56	
Hormone Receptor						
Unknown	N	146	73	52	21	0.0591
	Col (%)	47.87	41.01	59.09	53.85	
Negative	N	31	20	6	5	
	Col (%)	10.16	11.24	6.82	12.82	
Positive	N	128	85	30	13	
	Col (%)	41.97	47.75	34.09	33.33	
Year						
No	N	121	69	41	11	0.1421
	Col (%)	39.67	38.76	46.59	28.21	
Yes	N	184	109	47	28	
	Col (%)	60.33	61.24	53.41	71.79	
Grade						
Unknown	N	128	75	38	15	0.2340
	Col (%)	41.97	42.13	43.18	38.46	

		Surgical Choice				P-Value ₁
		ALL (N=305)	Breast Conserving Surgery (N=178)	Unilateral Mastectomy (N=88)	Bilateral Risk Reduction Mastectomy (N=39)	
Low	N	29	21	6	2	0.0591
	Col (%)	9.51	11.80	6.82	5.13	
Intermediate	N	80	50	22	8	
	Col (%)	26.23	28.09	25.00	20.51	
High	N	68	32	22	14	
	Col (%)	22.30	17.98	25.00	35.90	
Tumor Size						
Unknown	N	146	73	52	21	0.0591
	Col (%)	47.87	41.01	59.09	53.85	
Negative	N	31	20	6	5	
	Col (%)	10.16	11.24	6.82	12.82	
Positive	N	128	85	30	13	
	Col (%)	41.97	47.75	34.09	33.33	
Multicentric Disease						
No	N	246	167	51	28	<0.0001
	Col (%)	80.66	93.82	57.95	71.79	
Yes	N	59	11	37	11	
	Col (%)	19.34	6.18	42.05	28.21	

Footnote 1: P-Values obtained from univariate multinomial logistic regression with surgical choice as the response variable.

Modeling

Table 1b summarizes the results from the final multinomial logistic regression model with surgical choice as the response. Significant variables included age at diagnosis ($p=0.0076$), year of diagnosis ($p=0.0228$), tumor size ($p=0.0004$), multicentric disease ($p<0.0001$), and hormone receptor status ($p=0.0007$).

Table 1b: Predictors of Surgical Choice in Women with DCIS, $N=305$

Variable	P-Value	Group	Odds Ratio Estimate BCS	95% Confidence Interval BCS	Odds Ratio Estimate ULM	95% Confidence Interval ULM
Age	0.0076	< 50	Reference Group		Reference Group	
		50-64	3.26	(1.38, 7.68)	2.44	(0.99, 5.98)
		≥ 65	6.64	(2.02, 21.84)	3.43	(0.99, 11.83)
Year	0.0228	2005-2012	Reference Group		Reference Group	
		1997-2004	3.91	(1.48, 10.33)	2.51	(0.94, 6.71)
Tumor Size	0.0004	≥ 2	Reference Group		Reference Group	
		< 2	3.35	(1.24, 9.05)	0.67	(0.25, 1.77)
		Unknown	1.73	(0.56, 5.35)	0.39	(0.13, 1.19)
Multicentric Disease	<0.0001	Yes	Reference Group		Reference Group	
		No	6.69	(2.44, 18.32)	0.52	(0.22, 1.24)
Hormone Receptor Status	0.0007	Positive	Reference Group		Reference Group	
		Negative	0.69	(0.18, 2.62)	0.43	(0.10, 1.80)
		Unknown	0.189	(0.07, 0.50)	0.75	(0.28, 2.01)

When comparing breast conserving surgery to bilateral risk reduction mastectomy, several significant associations were found. In this multivariable model it was estimated that patients who were 50-64 years of age at diagnosis were 3.26 times more likely to choose breast conserving surgery over bilateral risk reduction mastectomy than patients who were less than 50 years old (95% CI 1.38-7.68). This effect is even larger when comparing patients 65 years and older who were 6.64 times more likely to choose breast conserving surgery as opposed to bilateral mastectomy than younger women. Patients

who were diagnosed between 1997 and 2004 were 3.91 times more likely to choose a breast conserving surgery over bilateral risk reduction mastectomy than patients who were diagnosed between 2005 and 2012. Those with small tumors (less than 2cm) were estimated to be 3.35 times more likely to choose breast conserving surgery over bilateral mastectomy than those with larger tumors (95% CI 1.24-9.05). Patients without a multicentric disease were estimated to be 6.69 times more likely to choose breast conserving surgery over bilateral mastectomy than patients with a multicentric disease. Also, those with a positive hormone receptor status were estimated to be 5.28 (1/0.189) times more likely to choose breast conserving surgery over bilateral mastectomy than patients with unknown hormone receptor status. When comparing the likeliness of choosing unilateral mastectomy to bilateral risk reduction mastectomy the differences weren't significant.

Hormone Therapy

Descriptive Statics

The sample size was reduced to 243 to only include patients who had breast conserving surgery or unilateral mastectomy, as these surgeries can be followed up with additional treatment. The sample was also reduced to patients who had at least 270 days of follow up to account for the time necessary for adjuvant treatment. It is interesting that more women with a multicentric disease choose not to undergo hormone therapy than to have this additional therapy (22.4% vs. 12.7%). Also, more women who chose unilateral mastectomy choose not to undergo hormone therapy as additional treatment (40.88% vs. 22.88%). Age, grade, tumor size, multicentric disease, hormone receptor status, radiation therapy, surgical choice, surgical margins, were potential predictors for the model (p-value<.2).

Table 2a: Demographics and Clinical Characteristics of Women with Stage 0 DCIS and Breast Conserving Surgery or Unilateral Mastectomy by Receipt of Hormone Therapy

		Hormone Therapy			P-Value ₁
		ALL (N=243)	YES (N=118)	NO (N=125)	
Age at Diagnosis					
<50	N	76	38	38	0.1525
	Col (%)	31.28	32.20	30.40	
50-64	N	100	54	46	
	Col (%)	41.15	45.76	36.80	
>=65	N	67	26	41	
	Col (%)	27.57	22.03	32.80	
Race					
Caucasian	N	135	66	69	0.4131
	Col (%)	55.56	55.93	55.20	
Hispanic	N	51	28	23	
	Col (%)	20.99	23.73	18.40	
Other	N	57	24	33	
	Col (%)	23.46	20.34	26.40	
Education					
High School or Less	N	117	59	58	0.5748
	Col (%)	48.15	50.00	46.40	
More Than High School	N	126	59	67	
	Col (%)	51.85	50.00	53.60	

		Hormone Therapy			P-Value ₁
		ALL (N=243)	YES (N=118)	NO (N=125)	
Insurance					
Other	N	2	1	1	0.3530
	Col (%)	0.82	0.85	0.80	
Managed	N	130	63	67	
	Col (%)	53.50	53.39	53.60	
Medicare	N	75	32	43	
	Col (%)	30.86	27.12	34.40	
Medicaid	N	36	22	14	
	Col (%)	14.81	18.64	11.20	
Menopausal Status					
Pre	N	85	42	43	0.8454
	Col (%)	34.98	35.59	34.40	
Post	N	158	76	82	
	(%)	65.02	64.41	65.60	
Hormone Receptor					
Unknown	N	116	58	58	0.0033
	Col (%)	47.74	49.15	46.40	
Negative	N	23	2	21	
	Col (%)	9.47	1.69	16.80	
Positive	N	104	58	46	
	Col (%)	42.80	49.15	36.80	

		Hormone Therapy			
		ALL (N=243)	YES (N=118)	NO (N=125)	
Year					
1997-2004	N	104	54	50	0.3645
	Col (%)	42.80	45.76	40.00	
2005-2012	N	139	64	75	
	Col (%)	57.20	54.24	60.00	
Grade					
Unknown	N	108	55	53	0.0345
	Col (%)	44.44	46.61	42.40	
Low	N	22	9	13	
	Col (%)	9.05	7.63	10.40	
Intermediate	N	66	39	27	
	Col (%)	27.16	33.05	21.60	
High	N	47	15	32	
	Col (%)	19.34	12.71	25.60	
Tumor Size					
Unknown	N	43	22	21	0.0689
	Col (%)	17.70	18.64	16.80	
<2	N	143	76	67	
	Col (%)	58.85	64.41	53.60	
≥2	N	57	20	37	
	Col (%)	23.46	16.95	29.60	

		Hormone Therapy			P-Value ₁
		ALL (N=243)	YES (N=118)	NO (N=125)	
Multicentric Disease					
No	N	200	103	97	0.0505
	Col (%)	82.30	87.29	77.60	
Yes	N	43	15	28	
	Col (%)	17.70	12.71	22.40	
Surgical Margins					
Unknown	N	125	44	81	<0.0001
	Col (%)	51.44	37.29	64.80	
Negative	N	100	60	40	
	Col (%)	41.15	50.85	32.00	
Positive	N	18	14	4	
	Col (%)	7.41	11.86	3.20	
Surgical Choice					
Breast Conserving Surgery	N	165	91	74	0.0031
	Col (%)	67.90	77.12	59.20	
Unilateral Mastectomy	N	78	27	51	
	Col (%)	32.10	22.88	40.80	
Radiation Therapy					
Yes	N	120	75	45	<0.0001
	(%)	49.38	63.56	36.00	
No	Col N	123	43	80	
	(%)	50.62	36.44	64.00	

Footnote 1: P-Values obtained from univariate logistic regression with hormone therapy as the response variable.

Modeling

Table 2b summarizes the results from modeling the probability of undergoing hormone therapy. The final model included hormone receptor status and the interaction between radiation therapy and age at diagnosis as explanatory variables.

Table 2b: Predictors of Receipt of Hormone Therapy for Women with DCIS and Breast Conserving Surgery or Unilateral Mastectomy

Variable	P-Value	Group		Odds Ratio Estimate	95% Confidence Interval
Hormone Receptor Status	0.0001	Negative		Reference Group	
		Positive		15.22	(3.25, 71.24)
		Unknown		12.07	(2.60, 56.05)
Radiation Therapy* Age at Diagnosis	0.0439	No RT	<50	Reference Group	
			50-64	3.35	(1.25, 8.99)
			>=65	0.99	(0.33, 2.99)
		Yes RT	<50	Reference Group	
			50-64	0.65	(0.26, 1.63)
			>=65	0.59	(0.21, 1.68)

The results showed that in this multivariable model a woman with a positive hormone receptor status was estimated to be 15.22 times more likely to undergo hormone therapy than a woman with a negative hormone receptor status (95% CI 3.25-71.24). Additionally a woman with an unknown hormone receptor status was likely to undergo hormone therapy than a patient with negative hormone receptor status (OR=12.07, 95% CI 2.60-56.05).

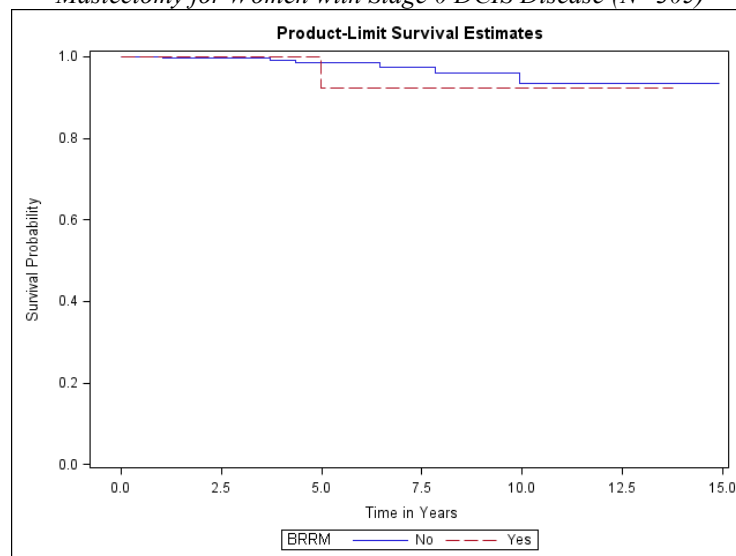
The interaction between age and radiation demonstrated that the relationship between age and likeliness of undergoing hormone therapy varies by receipt of radiation. For patients who haven't had

radiation therapy, the estimated odds of undergoing hormone therapy were 3.35 times higher for those younger than 50 years of age compared to those between 50 and 64 years old.

Survival Analysis

Figure 1 shows an overall survival probability plot for stratified by choice of surgery, bilateral risk reduction mastectomy or not.

Figure 1: Survival Probability Plot by Choice of Bilateral Risk Reduction Mastectomy for Women with Stage 0 DCIS Disease (N=305)



There was not a significant difference between the overall survival of women who received BRRM and those who did not (p-value=.5406). As there was no significant effect of whether or not a woman received a bilateral risk reduction mastectomy, it was not necessary to move on perform regression analysis of the survival data based on the Cox proportional hazards modelcontrolling for other covariates. Table 3 shows that only 7 women (2.3%) died from breast cancer.

Table 3: Survival Status for Women with DCIS Based on Surgical Choice

		Surgical Choice			
		ALL (N=305)	BCS (N=178)	ULM (N=88)	BRRM (N=39)
Alive	N	298	173	87	38
	Col (%)	97.70	97.19	98.86	97.44
Dead	N	7	5	1	1
	Col (%)	2.30	2.81	1.14	2.56

Conclusion

This analysis demonstrated that age at diagnosis, year, tumor size, presence of multicentric disease, and hormone receptor status were significant factors in predicting DCIS woman's likeliness surgical choice for breast conserving surgery compared to bilateral risk reduction mastectomy. It was shown that older women tend to be more likely to choose breast conserving surgery than older women. It is logical that older women were not going to want to undergo an extreme surgical operation either due to their health or desire to fight the disease. There was a similar relationship with women who were diagnosed between 1997 and 2004, those with small tumors (<2cm), those without multicentric disease, and those with a positive hormone receptor status. Breast cancer awareness has increased dramatically in recent years. Because of this, women diagnosed more recently appear to be more aware of their options so it's logical they were more likely to choose an extreme surgery. Women with larger tumors were also more willing to undergo bilateral risk reduction mastectomy. If a woman has more than one tumor (multicentric disease), then it makes sense that she would be more likely to take aggressive risk reduction actions. Women with unknown hormone receptor status may also want a bilateral mastectomy because they may not be candidates for hormone therapy after BCS or ULM, unfortunately the same could not be detected in the hormone negative women. When comparing likeliness of choosing a unilateral mastectomy to a bilateral risk reduction mastectomy, there were no significant comparisons. Choosing between a surgery that is less invasive and where the woman keeps her breasts as opposed to an extreme surgery where both breasts were removed is a drastic difference. However when a woman is choosing between invasive surgery where she loses just one breast or more invasive surgery where she loses two, the difference is not quite as extreme. Therefore it seems reasonable that only differences the between BCS and BRRM were significant.

Of the women who chose to undergo breast conserving surgery or a unilateral mastectomy, hormone receptor status and the interaction between age and radiation were significant predictors of receipt of hormone therapy. A patient with a positive or unknown hormone receptor status was much more likely to undergo hormone therapy than a patient with negative hormone receptor status. This result was not surprising, seeing as the hormone therapy is proven more effective for those with positive hormone receptors. It was interesting to note that the relationship between age and likeliness of undergoing hormone therapy varies between patients that have had radiation therapy and those who haven't. It was found that the estimated odds of undergoing hormone therapy were higher for patients who were in the mid-age range and who haven't had radiation therapy. This might be because women who are slightly older may be better positioned hormonally to try hormone therapy, but women who were older are not willing to undergo hormone therapy due to side effects or personal aversions.

Due to a small proportion of cases that resulted in death, no significant conclusions could be formed in examining the impact bilateral risk reduction mastectomies have on a patient's survival from breast cancer. DCIS is a noninvasive, stage 0 cancer, and treatment can prevent it from becoming a higher stage more invasive cancer. Because of this there were few deaths from DCIS in our cohort.

Further Research

Further research could be conducted in the area of survival analysis. DCIS has a relatively low death rate compared to other forms of breast cancer. There is however a higher risk of developing subsequent cancer in the other breast. Therefore disease free survival in which time until occurrence of subsequent breast cancer could be explored in addition to overall survival. This would examine the effectiveness of bilateral risk reduction mastectomies in preventing subsequent breast cancer. It is important to note however that disease free survival in breast cancer is complicated due to high lost to follow-up rates among breast cancer patients, especially in these healthier lower stage women.

References

Agresti, Alan. *An Introduction to Categorical Data Analysis*. New York: John Wiley & Sons, Inc. 1996.

"Breast Cancer Symptoms, Signs, Stages, Prevention, Detection, Facts, Treatment."
MedicineNet. N.p., n.d. Web. Nov. 2012.
<http://www.medicinenet.com/breast_cancer/article.htm>.

Chance, Beth. "Stat 418: Categorical Data Analysis, Lecture Notes." Cal Poly, San Luis Obispo. Lecture.

Cody, Ronald P. *Learning SAS by Example: A Programmer's Guide*. Cary, NC: SAS Institute, 2007. Print.

Ottesen, Rebecca. "Stat 330: Statistical Computing in SAS, Lecture Notes." Cal Poly, San Luis Obispo. Lecture.

"SAS Product Documentation." *SAS Product Documentation*. N.p., n.d. Web. Sept. 2012.
<<http://support.sas.com/documentation/>>.

Sklar, Jeff. "Stat 417: Survival Analysis, Lecture Notes." Cal Poly, San Luis Obispo. Lecture.

"WebMD Breast Cancer Guide." *WebMD*. WebMD, n.d. Web. Nov. 2012.
<<http://www.webmd.com/breast-cancer/guide/default.htm>>.

Appendix

The following SAS code was used to carry out this analysis:

(Please note that the data for this analysis is confidential and is on file with Rebecca Ottesen)

```
***SENIOR PROJECT: CREATING DATA SETS;
libname der 'C:\Users\Lauren\Dropbox\SrProj\NEW data\Derived';
libname raw 'C:\Users\Lauren\Dropbox\SrProj\NEW data\Raw';
libname myfmts 'C:\Users\Lauren\Dropbox\SrProj\NEW data';
options fmtsearch=(der.ddformats raw.formats myfmts); run;

/* viewing formats
proc format library=der.ddformats fmtlib;run;
proc format library=raw.formats fmtlib;run; */

*Create formats for grouping predictor variables;
proc format library=myfmts;
value agegroup
20-<50 = '<50'
50-<65 = '50-64'
65-100 = '>=65';
value racegroup
1 = 'Caucasian Non-Hispanic'
2 = 'Caucasian Hispanic'
3, 10 = 'African American and other Non-Hispanic'
5 = 'Asian, Pac Isl Non-Hispanic';
value racegrouppp
1 = 'Caucasian Non-Hispanic'
2 = 'Hispanic'
3, 5, 10 = 'Other Non-Hispanic';
value racegroup_w
1 = 'Caucasian Non-Hispanic'
2, 3, 5, 10 = 'Other';
value edugroup
-1-3 = 'High School or Less'
4-7 = 'Higher than High School';
value yeargroup
1997-2004='1997-2004'
2005-2012='2005-2012';
value insgroup
2, 3, 5, 5.5, 5.75 = 'Medicare'
4 = 'Medicaid'
1 = 'Managed'
-1, 6, 7, 0 = 'Other';
value tumorgroup
0-2 = '<2'
2-11 = '>=2'
-1 = 'Unknown';
value margingroup
-1 = 'Unknown'
0 = 'Negative'
3, 4 = 'Positive';
value $outgroup
'BCS' = 'No BRRM'
'ULM' = 'No BRRM'
```

```

'BRRM' = 'BRRM';
value adjTX
2, 99 = 'No'
4 = 'Yes';
run;

**CREATE COHORT;

*Merge solid tumor stage and clinical characteristics data sets;
data stage_clinical;
merge raw.solid_tumor_stage der.clinical_characteristics;
by pid dxid tumorid;
run;

*Merge above with patient characteristics to create the cohort;
data cohort;
merge stage_clinical der.patient_characteristics;
by pid dxid;
if histology ^= 2 then delete;
if fu90 ^= 1 then delete;
if stage_final ^= 22 then delete;
run;

**BREAST CONSERVING SURGERY PATIENTS;

*Find BCS patients;
data bcs; set der.surgical_information;
where dsgroup=2;
run;

*Match to those in cohort;
data bcs_dcis;
merge bcs(in=inbcs keep=pid dxid tumorid dsgroup) cohort(in=incohort);
by pid dxid tumorid;
if inbcs and incohort;
surrgroup="BCS";
run;
*178 patients (with 90 day followup);

**MASTECTOMY PATIENTS;

*Find Mastectomy patients;
data mast; set der.surgical_information;
where dsgroup=3;
run;

*Match to those in cohort;
data mast_dcis;
merge mast(in=inmast keep=pid dxid tumorid dsgroup) cohort(in=incohort);
by pid dxid tumorid;
if inmast and incohort;
run;
*127 patients (with 90 day followup);

```

```

*Separate Mastectomy and Radiation Therapy patients from Treatment data set;
data treatment radiation; set raw.treatment;
if txcat=11 and proctype in(11 12 14 15 16) then output treatment;
else if txcat=16 and proctype=1 then output treatment;
else if txcat=10 and indication in(1, 2, 4) then output radiation;
run;

*Find Contralateral Mastectomy patients;
data rawtrt_sts;
merge treatment(in=intrt) raw.solid_tumor_stage;
by pid dxid;
if intrt then do;
if side^=laterality then flag2="contralateral";
else delete;
end;
else delete;
run;

*Merge Contralateral Mastectomy patients with all Mastectomy patients;
data mast2;
merge rawtrt_sts(keep= pid dxid flag2) mast_dcis(in=inmast);
by pid dxid;
if inmast;
run;

**MERGE PATIENTS WITH OTHER DATA SETS TO OBTAIN PREDICTOR VARIABLES;

*Merge bcs with study accession by pid;
data bcs_dcis2;
merge bcs_dcis(in=inbcs) raw.study_accession;
by pid;
if inbcs;
run;

*Merge mast with study accession by pid;
data mast_dcis2;
merge mast2(in=inmast) raw.study_accession;
by pid;
if inmast;
run;

*Merge bcs with adjuvant drug therapy by pid, dxid;
data bcs_dcis3;
merge bcs_dcis2(in=inbcs) der.adjuvant_drug_therapy;
by pid dxid;
if inbcs;
run;

*Merge mast with adjuvant drug therapy by pid, dxid;
data mast_dcis3;
merge mast_dcis2(in=inmast) der.adjuvant_drug_therapy;
by pid dxid;
if inmast;
run;

```

```

*Merge bcs with surgical information and breast diagnosis by pid, dxid,
tumorid;
data bcs_dcis4;
merge bcs_dcis3(in=inbcs) der.surgical_information raw.breast_diagnosis;
by pid dxid tumorid;
if inbcs;
run;

*Merge mast with surgical information and breastdx by pid, dxid, tumorid;
data mast_dcis4;
merge mast_dcis3(in=inmast) der.surgical_information raw.breast_diagnosis;
by pid dxid tumorid;
if inmast;
run;

**CREATE DATA SET FOR OUTCOME ANALYSIS;

data der.allpatients;
set bcs_dcis4(in=inbcs) mast_dcis4;
length outcome $4;
*Determine outcome;
if inbcs then outcome="BCS";
else if flag2="contralateral" then outcome="BRRM";
else if flag2=" " then outcome="ULM";
*Create year to use as a predictor variable;
year=year(tumordxdt);
if dcispathtumsize= . then dcispathtumsize=-1;
*Assign user created formats to variables;
format agedx agegroup.
race_eth racegroup.
edustat edugroup.
year yeargroup.
insurance insgroup.
dcispathtumsize tumorgroup.;
*Keep only variables used in analysis;
keep pid agedx race_eth edustat year dcishistogrds insurance dcispathtumsize
dcismulcent menopause HR outcome fu90 fu270 adjtxgroup;
run;
*305 patients (90 day followup);

*CREATE DATA SET FOR HORMONE THERAPY ANALYSIS;

*Merge radiation therapy patients with solid tumor stage data set;
data radiation2;
merge radiation(in=inr) raw.solid_tumor_stage;
by pid dxid;
if inr;
run;

*Merge above data with patient characteristics;
data radiation3;
merge radiation2(in=inr) der.patient_characteristics;
by pid;
if inr;
if side^=finallat then delete;

```

```

run;

*Match radiation therapy patients to those in cohort;
data radiation_cohort;
merge radiation3(in=inr) cohort(in=inc);
by pid;
if inr and inc;
where fu270=1;
run;

*Merge with all patients data set to create final data set for hormone
therapy analysis;
proc sort data=der.allpatients; by pid; run;
data der.hormone;
merge der.allpatients(in=inall) radiation_cohort;
by pid;
if inall;
*Eliminate BRRM patients;
if outcome in('BCS','ULM');
*Use only patients with 270 day follow-up rule;
where fu270=1;
if txcat=10 then radiation='Yes';
else radiation='No';
if surgmargin=. then surgmargin=-1;
*Assign format to adjuvant treatment group variable;
format adjtxgroup adjtx.
surgmargin margingroup.;
if dcispathtumsize= . then dcispathtumsize=-1;
*Keep only variables used in analysis;
keep pid agedx race_eth edustat year dcishistogrd insurance dcispathtumsize
dcismulcent menopause HR outcome surgmargin radiation adjtxgroup;
run;
*243 Patients (270 day followup);

*CREATE DATA SET FOR SURVIVAL ANALYSIS;
proc contents data=bcs_dcis4; run;
proc contents data=mast_dcis4; run;

data der.survival;
set bcs_dcis4(in=inbcs) mast_dcis4;
length outcome $4 BRRM $3.;
*Determine outcome;
if inbcs then outcome="BCS";
else if flag2="contralateral" then outcome="BRRM";
else if flag2=" " then outcome="ULM";
if outcome='BCS' then BRRM='No';
else if outcome='ULM' then BRRM='No';
else if outcome='BRRM' then BRRM='Yes';
*Create year to use as a predictor variable;
year=year(tumordxdt);
time=osdt-tumordxdt;
time_yr=time/365;
*Assign user created formats to variables;
format agedx agegroup.

```

```

race_eth racegroup.
edustat edugroup.
year yeargroup.
insurance insgroup.
dcispathtumsize tumorgroup.;
*Keep only variables used in analysis;
keep pid agedx race_eth edustat year dcishistogrd insurance dcispathtumsize
dcismulcent menopause HR outcome fu90 fu270 adjtxgroup event osdt ossource
BRRM time time_yr;
run;

***ANALYSIS FOR SURGICAL CHOICE;

**FREQUENCIES;

*AGE;
proc freq data=der.allpatients;
tables outcome*agedx;
run;

*RACE;
proc freq data=der.allpatients;
tables outcome*race_eth;
run;

*EDUCATION;
proc freq data=der.allpatients;
tables outcome*edustat;
run;

*YEAR;
proc freq data=der.allpatients;
tables outcome*year;
run;

*GRADE;
proc freq data=der.allpatients;
tables outcome*dcishistogrd;
run;

*INSURANCE;
proc freq data=der.allpatients;
tables outcome*insurance;
run;

*TUMOR SIZE;
proc freq data=der.allpatients;
tables outcome*dcispathtumsize;
run;

*MULTICENTRIC DISEASE;
proc freq data=der.allpatients;
tables outcome*dcismulcent;
run;

*MENOPAUSAL STATUS;
proc freq data=der.allpatients;

```



```

tables outcome*menopause;
run;

*HORMONE RECEPTOR STATUS;
proc freq data=der.allpatients;
tables outcome*HR;
run;

**UNIVARIATE RELATIONSHIPS BETWEEN POTENTIAL PREDICTORS AND OUTCOME;

*AGE;
proc logistic data=der.allpatients;
class agedx/param=ref;
model outcome=agedx/link=glogit;
run;
*p-value = 0.0037;

*RACE;
proc logistic data=der.allpatients;
class race_eth/param=ref;
model outcome=race_eth/link=glogit;
run;
*p-value = 0.3448;

*EDUCATION;
proc logistic data=der.allpatients;
class edustat/param=ref;
model outcome=edustat/link=glogit;
run;
*p-value = 0.0283;

*YEAR;
proc logistic data=der.allpatients;
class year/param=ref;
model outcome=year/link=glogit;
run;
*p-value = 0.1421;

*GRADE;
proc logistic data=der.allpatients;
class dcishistogr/param=ref;
model outcome=dcishistogr/link=glogit;
run;
*p-value = 0.2340;

*INSURANCE;
proc logistic data=der.allpatients;
class insurance/param=ref;
model outcome=insurance/link=glogit;
run;
*p-value = 0.0758;

*TUMOR SIZE;
proc logistic data=der.allpatients;
class dcispathtumsize/param=ref missing;
model outcome=dcispathtumsize/link=glogit;

```

```

run;
*p-value = 0.0002;

*MULTICENTRIC DISEASE;
proc logistic data=der.allpatients;
class dcismulcent/param=ref;
model outcome=dcismulcent/link=glogit;
run;
*p-value < 0.0001;

*MENOPAUSAL STATUS;
proc logistic data=der.allpatients;
class menopause/param=ref;
model outcome=menopause/link=glogit;
run;
*p-value = 0.0012;

*HORMONE RECEPTOR STATUS;
proc logistic data=der.allpatients;
class HR/param=ref;
model outcome=HR/link=glogit;
run;
*p-value = 0.0591;

**NARROWED DOWN POTENTIAL PREDICTORS (p-value<.2):
Age, Education, Year, Insurance, Tumor Size, Multicentric Disease, Menopausal
Status, Hormone Receptor Status;

**MODELING WITH POTENTIAL PREDICTORS;

*Model with narrowed down potential predictors;
proc logistic data=der.allpatients;
class agedx edustat year insurance dcispathtumsize dcismulcent menopause
HR/param=ref missing;
model outcome=agedx edustat year insurance dcispathtumsize dcismulcent
menopause HR/link=glogit;
run;

*Drop Insurance;
proc logistic data=der.allpatients;
class agedx edustat year dcispathtumsize dcismulcent menopause HR/param=ref
missing;
model outcome=agedx edustat year dcispathtumsize dcismulcent menopause
HR/link=glogit;
run;

*Drop Menopausal Status;
proc logistic data=der.allpatients;
class agedx edustat year dcispathtumsize dcismulcent HR/param=ref missing;
model outcome=agedx edustat year dcispathtumsize dcismulcent HR/link=glogit;
run;

*Drop Education;
proc logistic data=der.allpatients;
class agedx year dcispathtumsize dcismulcent HR/param=ref missing;

```

```

model outcome=agedx year dcispathtumsize dcismulcent HR/link=glogit;
run;

**MODELING WITH INTERACTIONS;

*Interaction of Age and Race;
proc logistic data=der.allpatients;
class agedx year dcispathtumsize HR dcismulcent/param=ref missing;
model outcome=agedx year dcispathtumsize HR dcismulcent
agedx*race_eth/link=glogit;
run;

*Take out Race, Interaction between Age and Year;
proc logistic data=der.allpatients;
class agedx year dcispathtumsize HR dcismulcent/param=ref missing;
model outcome=agedx year dcispathtumsize HR dcismulcent
agedx*year/link=glogit;
run;

*Interaction between Age and Tumor Size;
proc logistic data=der.allpatients;
class agedx year dcispathtumsize HR dcismulcent/param=ref missing;
model outcome=agedx year dcispathtumsize HR dcismulcent
agedx*dcispathtumsize/link=glogit;
run;

*Interaction between Age and Hormone Receptor Status;
proc logistic data=der.allpatients;
class agedx dcispathtumsize HR dcismulcent/param=ref missing;
model outcome=agedx dcispathtumsize HR dcismulcent agedx*HR/link=glogit;
run;

*Interaction between Age and Multicentric Disease;
proc logistic data=der.allpatients;
class agedx dcispathtumsize HR dcismulcent/param=ref missing;
model outcome=agedx dcispathtumsize HR dcismulcent
agedx*dcismulcent/link=glogit;
run;

*Interaction between Age and Education;
proc logistic data=der.allpatients;
class agedx edustat dcispathtumsize HR dcismulcent/param=ref missing;
model outcome=agedx edustat dcispathtumsize HR dcismulcent
agedx*edustat/link=glogit;
run;

**BEST MODEL: AGE, YEAR, TUMOR SIZE, MULTICENTRIC DISEASE, HORMONE RECEPTOR
STATUS;
proc logistic data=der.allpatients;
class agedx(ref='<50') year(ref='2005-2012') dcispathtumsize(ref='>=2')
dcismulcent HR(ref='Negative')/param=ref;
model outcome(ref='BRRM')=agedx year dcispathtumsize dcismulcent
HR/link=glogit aggregate scale=D influence rsquare stb iplots;
run;

```

```

**CHECK OTHER MODEL ASSUMPTIONS;
ods graphics on;
proc corr data=der.allpatients spearman;
var agedx year dcispathtumsize dcismulcent HR;
run;
proc genmod data=der.allpatients plots=(stdreschi reschi leverage dobs);
class outcome agedx year dcispathtumsize dcismulcent HR/ param=ref;
model outcome=agedx year dcispathtumsize dcismulcent HR/ link=cumcll
dist=multinomial;
output out=resids reschi=pearson stdreschi=sdtpearson leverage=lev
cooksd=infl;
run;
ods graphics off;

```

```

***ANALYSIS FOR HORMONE THERAPY;

```

```

**FREQUENCIES;

```

```

*AGE;

```

```

proc freq data=der.hormone;
tables adjtxgroup*agedx;
run;

```

```

*RACE;

```

```

proc freq data=der.hormone;
tables adjtxgroup*race_eth;
run;

```

```

*EDUCATION;

```

```

proc freq data=der.hormone;
tables adjtxgroup*edustat;
run;

```

```

*YEAR;

```

```

proc freq data=der.hormone;
tables adjtxgroup*year;
run;

```

```

*GRADE;

```

```

proc freq data=der.hormone;
tables adjtxgroup*dcishistogr;
run;

```

```

*INSURANCE;

```

```

proc freq data=der.hormone;
tables adjtxgroup*dcishistogr;
run;

```

```

*TUMOR SIZE;

```

```

proc freq data=der.hormone;
tables adjtxgroup*dcispathtumsize;
run;

```

```

*MULTICENTRIC DISEASE;

```

```

proc freq data=der.hormone;
tables adjtxgroup*dcismulcent;
run;

*MENOPAUSAL STATUS;
proc freq data=der.hormone;
tables adjtxgroup*menopause;
run;

*HORMONE RECEPTOR STATUS;
proc freq data=der.hormone;
tables adjtxgroup*HR;
run;

*OUTCOME;
proc freq data=der.hormone;
tables adjtxgroup*outcome;
run;

*SURGICAL MARGINS;
proc freq data=der.hormone;
tables adjtxgroup*surgmargin;
run;

*RADIATION THERAPY;
proc freq data=der.hormone;
tables adjtxgroup*radiation;
run;

**UNIVARIATE RELATIONSHIPS BETWEEN POTENTIAL PREDICTORS AND ADJUVANT
TREATMENT GROUP;

*AGE;
proc logistic data=der.hormone;
class agedx/param=ref;
model adjtxgroup=agedx;
run;
*p-value = 0.1525;

*RACE;
proc logistic data=der.hormone;
class race_eth/param=ref;
model adjtxgroup=race_eth;
run;
*p-value = 0.4131;

*EDUCATION;
proc logistic data=der.hormone;
class edustat/param=ref;
model adjtxgroup=edustat;
run;
*p-value = 0.5748;

*YEAR;
proc logistic data=der.hormone;
class year/param=ref;

```

```

model adjtxgroup=year;
run;
*p-value = 0.3645;

*GRADE;
proc logistic data=der.hormone;
class dcishistogrd/param=ref;
model adjtxgroup=dcishistogrd;
run;
*p-value = 0.0345;

*INSURANCE;
proc logistic data=der.hormone;
class insurance/param=ref;
model adjtxgroup=insurance;
run;
*p-value = 0.3530;

*TUMOR SIZE;
proc logistic data=der.hormone;
class dcispathtumsize/param=ref;
model adjtxgroup=dcispathtumsize;
run;
*p-value = 0.0689;

*MULTICENTRIC DISEASE;
proc logistic data=der.hormone;
class dcismulcent/param=ref;
model adjtxgroup=dcismulcent;
run;
*p-value = 0.0505;

*MENOPAUSAL STATUS;
proc logistic data=der.hormone;
class menopause/param=ref;
model adjtxgroup=menopause;
run;
*p-value = 0.8454;

*HORMONE RECEPTOR STATUS;
proc logistic data=der.hormone;
class HR/param=ref;
model adjtxgroup=HR;
run;
*p-value = 0.0033;

*OUTCOME;
proc logistic data=der.hormone;
class outcome/param=ref;
model adjtxgroup=outcome;
run;
*p-value = 0.0031;

*SURGICAL MARGINS;
proc logistic data=der.hormone;
class surgmargin/param=ref;
model adjtxgroup=surgmargin;

```

```

run;
*p-value < 0.0001;

*RADIATION THERAPY;
proc logistic data=der.hormone;
class radiation/param=ref;
model adjtxgroup=radiation;
run;
*p-value < 0.0001;

**NARROWED DOWN POTENTIAL PREDICTORS (p-value<.2):
Age, Grade, Tumor Size, Multicentric Disease, HR, Outcome, Surgical Margins,
Radiation Therapy;

**MODELING WITH POTENTIAL PREDICTORS;

*Model with narrowed down potential predictors;
proc logistic data=der.hormone;
class agedx dcishistogrd dcispathtumsize dcismulcent HR outcome surgmargin
radiation/param=ref;
model adjtxgroup=agedx dcishistogrd dcispathtumsize dcismulcent HR outcome
surgmargin radiation/lackfit;
run;

*Drop Radiation;
proc logistic data=der.hormone;
class agedx dcishistogrd dcispathtumsize dcismulcent HR outcome
surgmargin/param=ref;
model adjtxgroup=agedx dcishistogrd dcispathtumsize dcismulcent HR outcome
surgmargin/lackfit;
run;

*Drop Outcome;
proc logistic data=der.hormone;
class agedx dcishistogrd dcispathtumsize dcismulcent HR surgmargin/param=ref;
model adjtxgroup=agedx dcishistogrd dcispathtumsize dcismulcent HR
surgmargin/lackfit;
run;

*Drop Tumor Size;
proc logistic data=der.hormone;
class agedx dcishistogrd dcismulcent HR surgmargin/param=ref;
model adjtxgroup=agedx dcishistogrd dcismulcent HR surgmargin/lackfit;
run;

*Drop Multicentric Disease;
proc logistic data=der.hormone;
class agedx dcishistogrd HR surgmargin/param=ref;
model adjtxgroup=agedx dcishistogrd HR surgmargin/lackfit;
run;

*Drop Grade;
proc logistic data=der.hormone;
class agedx HR surgmargin/param=ref;
model adjtxgroup=agedx HR surgmargin/lackfit;

```

```

run;

*Drop Age;
proc logistic data=der.hormone;
class HR surgmargin/param=ref;
model adjtxgroup=HR surgmargin/lackfit;
run;

**INTERACTIONS WITH AGE;

*Interaction between Age and HR;
proc logistic data=der.hormone;
class agedx HR surgmargin/param=ref;
model adjtxgroup=agedx HR surgmargin agedx*HR/lackfit;
run;

*Interaction between Age and Surgical Margins;
proc logistic data=der.hormone;
class agedx HR surgmargin/param=ref;
model adjtxgroup=agedx HR surgmargin agedx*surgmargin/lackfit;
run;

*Interaction between Age and Radiation without surgical margins;
proc logistic data=der.hormone;
class agedx HR radiation/param=ref;
model adjtxgroup=agedx HR radiation agedx*radiation/lackfit;
run;

*Interaction between Age and Grade;
proc logistic data=der.hormone;
class agedx dcishistogr dcispathtumsize HR surgmargin/param=ref;
model adjtxgroup=agedx dcishistogr dcispathtumsize HR surgmargin
agedx*dcishistogr/lackfit;
run;

*Interaction between Age and Tumor Size;
proc logistic data=der.hormone;
class agedx dcishistogr dcispathtumsize HR surgmargin/param=ref;
model adjtxgroup=agedx dcishistogr dcispathtumsize HR surgmargin
agedx*dcispathtumsize/lackfit;
run;

**BEST MODEL: HORMONE RECEPTOR STATUS AND AGE-RADIATION INTERACTION;
proc logistic data=der.hormone;
class agedx(ref='<50') HR(ref='Negative') radiation/param=ref;
model adjtxgroup(ref='No')=agedx HR radiation agedx*radiation/lackfit
aggregate scale=D influence rsquare stb iplots;
oddsratio agedx / diff=ref;
oddsratio hr / diff=ref;
run;

**CHECK OTHER MODEL ASSUMPTIONS;
ods graphics on;
proc genmod data=der.hormone plots=(stdreschi reschi leverage dobs);
class adjtxgroup agedx HR radiation/param=ref;

```



```

model adjtxgroup=agedx HR radiation agedx*radiation/link=logit dist=bin;
output out=resids reschi=pearson stdreschi=stdpearson leverage=lev
cooksd=infl;
run;
proc corr data=der.hormone spearman;
var agedx HR;
run;
ods graphics off;

***ANALYSIS FOR SURVIVAL;

proc freq data=der.survival;
tables outcome*event;
run;

***outcome and all covariates;
proc phreg data=der.survival plots(overlay)=survival;
class outcome agedx race_eth edustat year dcishistogrd insurance
dcispathtumsize dcismulcent menopause HR;
model time_yr*event(0)=outcome agedx race_eth edustat year dcishistogrd
insurance dcispathtumsize dcismulcent menopause HR;
run;

*remove HR, dcishistogrd, and year;
proc phreg data=der.survival plots(overlay)=survival;
class outcome agedx race_eth edustat insurance dcispathtumsize dcismulcent
menopause;
model time_yr*event(0)=outcome agedx race_eth edustat insurance
dcispathtumsize dcismulcent menopause;
run;

*remove tumor size, insurance, edustat;
proc phreg data=der.survival plots(overlay)=survival;
class outcome agedx race_eth dcismulcent menopause;
model time_yr*event(0)=outcome agedx race_eth dcismulcent menopause;
run;

*remove multicentric and race;
proc phreg data=der.survival plots(overlay)=survival;
class outcome agedx menopause;
model time_yr*event(0)=outcome agedx menopause;
run;

***Model with BRRM vs No BRRM;
proc phreg data=der.survival plots(overlay)=survival;
class BRRM agedx race_eth edustat year dcishistogrd insurance dcispathtumsize
dcismulcent menopause HR;
model time_yr*event(0)=BRRM agedx race_eth edustat year dcishistogrd
insurance dcispathtumsize dcismulcent menopause HR;
run;

*remove HR, year, grade, tumor size;
proc phreg data=der.survival plots(overlay)=survival;
class BRRM agedx race_eth edustat insurance dcismulcent menopause;

```

```

model time_yr*event(0)=BRRM agedx race_eth edustat insurance dcismulcent
menopause;
run;

*remove insurance and race;
proc phreg data=der.survival plots(overlay)=survival;
class BRRM agedx edustat dcismulcent menopause;
model time_yr*event(0)=BRRM agedx edustat dcismulcent menopause;
run;

ods graphics on;
proc lifetest data=der.survival plots=survival (nocensor);
strata BRRM;
time time_yr*event(0);
test agedx race_eth edustat year dcishistogrd insurance dcispathtumsize
dcismulcent menopause HR;
label time_yr='Time in Years';
run;
ods graphics off;

```